

# **Artificial Intelligent System for Protein Superfamily Classification**

## **FIELD OF THE INVENTION**

The invention is related to an artificial intelligent (abbreviated as AI) system for  
5 protein superfamily classification, especially to an AI system combined with the fuzzy  
logic system.

## **BACKGROUND OF THE INVENTION**

In bioinformatics technology, a classification, such as a protein superfamily  
classification, is an important task and costs more time and expenses. In recent years,  
10 neural network (abbreviated as NN) technology is widely used in analysis of  
bioinformatics.

Several research works have shown that NN technology can be used in biology  
chemistry family classification. For example, US pattern No. 5845049 has proposed a  
molecule sequencing method using NN technology.

15 Since the main coding method is N-GRAM, the amount of data and computation is  
quite large, hence high-end computers usually perform the classification process.  
Moreover, the accuracy of NN-based algorithms is not enough, and the efficiency of  
performing classification on computers is also not good. As a result, both drawbacks  
limit the applicability of the NN-based approaches.

20

## **SUMMARY OF THE INVENTION**

The invention proposes an AI system for protein family classification, uses the fuzzy  
logic theory in an NN system, and improves robustness, convergence and correctness by

utilizing the memory and learning characteristics of NN systems, the determination expertise of the fuzzy theory which introduced the so called expert knowledge, and a content addressable memory (abbreviated as CAM) concept used to speedup input vector encoding, so that the hardware of the algorithm can work faster.

5

## **BRIEF DESCRIPTION OF THE DRAWINGS**

Fig. 1 shows the architecture of the invention.

Fig. 2A shows the search process of a traditional search approach.

Fig. 2B shows the search process of CAM.

10      Fig. 3A shows the first example of the combinations of a fuzzy logic system and a NN system.

Fig. 3B shows the second example of the combinations of a fuzzy logic system and a NN system.

15      Fig. 3C shows the third example of the combinations of a fuzzy logic system and a NN system.

## **DESCRIPTION OF THE PREFERRED EMBODIMENT**

The invention proposes an AI system for protein superfamily classification, which is an expert system utilizing NN technology and the fuzzy logic system. The expert system

can organize the experts' knowledge and simulate the inference behavior of experts, to classify a protein family.

First, the experts' knowledge consists of linguistic variables and a fuzzy set, and a fuzzy expert system is built by the derived linguistic variables and fuzzy set. The inference process of the fuzzy logic can be represented by a resolution function. Then, various algorithms in NN are used to adapt the parameters of the fuzzy expert system. The fuzzy expert system automatically updates its knowledge base, hence the fuzzy inference engine works correctly as time goes by.

The proposed system is used to improve the efficiency of the protein family (e.g., protein super family) classification. Fig. 1 shows the architecture of the proposed system. The AI system integrates a fuzzy logic system into an NN system to classify the protein super family sequence.

There are various combinations of a fuzzy logic system and an NN system. Fig. 3A shows the first example of the combinations. The input data  $X_1 \sim X_n$  are processed by a fuzzy set  $A_i$ . Then the results are classified by membership functions  $\mu_{A1} \sim \mu_{An}$  and the aggregation operator  $\otimes$  to obtain the classification result  $Y$ . Fig. 3B shows the second example of the combinations. It directly codes the fuzzy logic system into the NN system. The input data  $X_1 \sim X_n$  are processed by a fuzzy set  $A_i$  to obtain  $Y = X_1 \otimes X_2 \otimes \dots$ . Fig. 3C shows the third example of the combinations. Multiple input  $X_i$ s are processed by a fuzzy transfer relation  $R$  (e.g., t-norm) to obtain the result  $Y$ .

In addition, CAM concept is used in the hardware architecture to make the search process faster. It also reduces the size of the hardware architecture so that the hardware can be designed as a commercialized interface card.

Fig. 2A and 2B show the search processes of a traditional approach and CAM, respectively. In a traditional computer-based search method, the address to be searched is inputted (Step 201), and personal computers or other computation devices then search the address-content table 202 to obtain the corresponding content (Step 203) and  
5 compare the content (Step 204). The efficiency of the traditional approach is low, since it searches the address-content table sequentially.

In CAM, after the content is inputted, the result can be obtained by applying logical operations (Step 213) to the address-content table 212, hence the search-efficiency is improved.

10 The proposed AI system integrates the fuzzy inference theory into an NN system, and improves robustness, convergence and correctness by utilizing the memory and learning characteristics of NN systems, the determination expertise of the fuzzy inference theory, and a content addressable memory to make the system can be commercialized easily.

15 While the preferred embodiment of the invention has been set forth for the purpose of disclosure, modifications of the disclosed embodiment of the invention as well as other embodiments thereof may occur to those skilled in the art. Accordingly, the appended claims are intended to cover all embodiments, not departing from the spirit and scope of the invention.

20 While the preferred embodiment of the invention has been set forth for the purpose of disclosure, modifications of the disclosed embodiment of the invention as well as other embodiments thereof may occur to those skilled in the art. Accordingly, the appended claims are intended to cover all embodiments which do not depart from the spirit and scope of the invention.